BANA 4090: Chapter 4: ARIMA models (Part III)

Zhaohu(Jonathan) Fan

07/13/2021

- Main topics:
  - ARIMA models
  - Backshift notation

ARIMA models

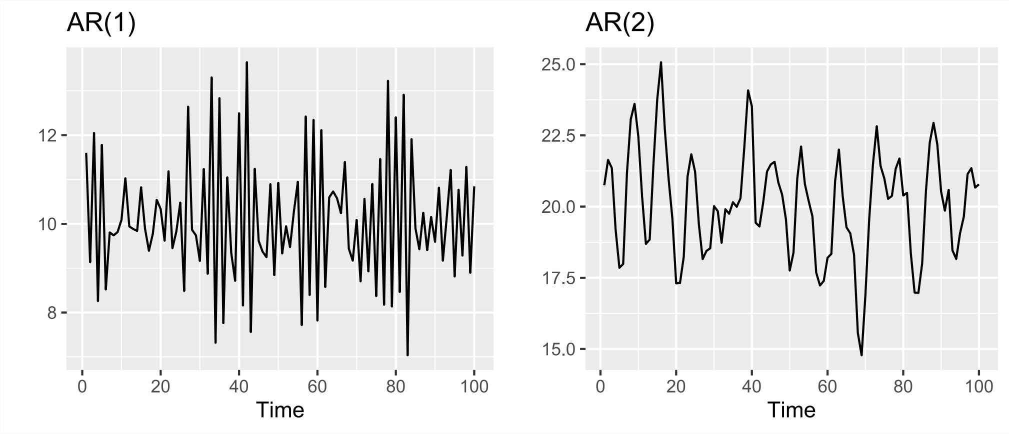
ARIMA models

- Autoregressive Integrated Moving Average models ($\mathbf{ARIMA}\ (p,d,q)\ \mathbf{model}$):
  - $\mathbf{AR}$: $p =$ order of the autoregressive part.
  - $\mathbf{I}$: $d =$ degree of first differencing involved.
  - $\mathbf{MA}$: $q =$ order of the moving average part.
- Examples:
  - White noise model: ARIMA(0,0,0)
  - Random walk: ARIMA(0,1,0) with no constant
  - Random walk with drift: ARIMA(0,1,0) with with no constant
  - AR (p): ARIMA $(p, 0, 0)$
  - MA (q): ARIMA $(0, 0, q)$

Autoregressive models (AR)

- Autoregressive (AR) models: $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$, where ε_t is white noise. This is a multiple regression with **lagged values** of y_t as predictors.

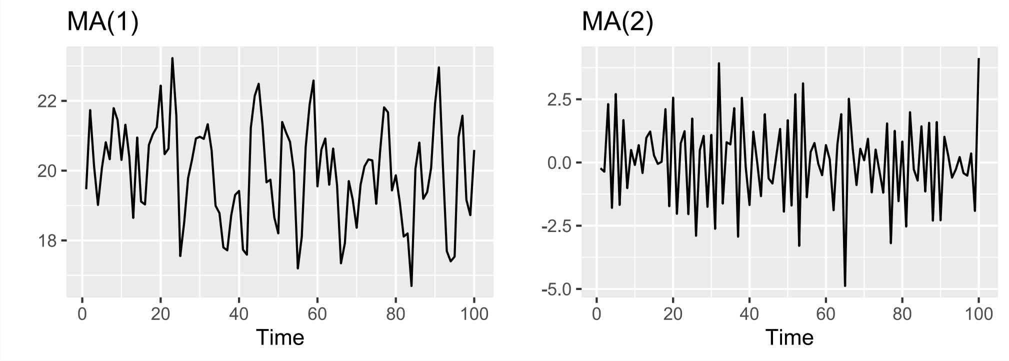
```
set.seed(1)
p1 <- autoplot(10 + arima.sim(list(ar = -0.8), n = 100)) +
  ylab("") + ggtitle("AR(1)")
p2 <- autoplot(20 + arima.sim(list(ar = c(1.3, -0.7)), n = 100)) +
  ylab("") + ggtitle("AR(2)")
gridExtra::grid.arrange(p1,p2,nrow=1)
```



Moving Average models (MA)

- Moving Average (MA) models: $y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$,where $\varepsilon_t$is white noise.
- This is a multiple regression with **past errors** as predictors. **Don't confuse this with moving average smoothing!**

```
set.seed(2)
p1 <- autoplot(20 + arima.sim(list(ma = 0.8), n = 100)) +
  ylab("") + ggtitle("MA(1)")
p2 <- autoplot(arima.sim(list(ma = c(-1, +0.8)), n = 100)) +
  ylab("") + ggtitle("MA(2)")
gridExtra::grid.arrange(p1,p2,nrow=1)
```



ARIMA models

- Autoregressive Moving Average models **(ARMA)**:
  - $y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$
- Predictors include both **lagged values of y_t and lagged errors.**
- Conditions on coefficients ensure invertibility.
- Autoregressive Integrated Moving Average models **(ARIMA)**:
  - Combine ARMA model with **differencing**.
  - $(1 - B)^d y_t$ follows an ARMA model.

Backshift notation

Backshift notation

- A very useful notational device is the backward shift operator, B, which is used as follows: $By_t = y_{t-1}$.
- In other words, B, operating on y_t, has the effect of **shifting the data back one period**.
- Two applications of B to y_t **shifts the data back two periods**: $B(By_t) = B^2 y_t = y_{t-2}$.
- For monthly data, if we wish to shift attention to the same month last year, then B^{12} is used, and the notation is $B^{12} y_t$ = y_{t-12}.

Backshift notation (cont'd)

- The backward shift operator is convenient for describing the process of **differencing**.
- A first difference can be written as $y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t$.
  - Note: a first difference is represented by $(1 - B)$.
- Similarly, if second-order differences (i.e.,first differences of first differences) have to be computed, then: $y''_t = y_t - 2y_{t-1} + y_{t-2} = (1 - B)^2 y_t$.

Backshift notation (cont'd)

- Second-order difference is denoted $(1 - B)^2$.
- **Second-order difference** is not the same as a **second difference**, which would be denoted $1 - B^2$;
- In general, a dth-order difference can be written as

$$(1 - B)^d y_t.$$

- A seasonal difference followed by a first difference can be written as

$$(1 - B)(1 - B^m) y_t.$$

- Note:
  - A seasonal difference is the difference between an observation and the corresponding observation from the previous year.
  - $y'_t = y_t - y_{t-m}$, where m= number of seasons.
  - For monthly data m=12.
  - For quarterly data m=4.